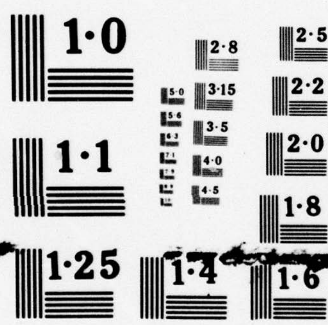
AD-A052 596 ILLINOIS UNIV AT CHICAGO CIRCLE DEPT OF MATHEMATICS F/G 12/1
SAMPLING DESIGNS WITH REDUCED SUPPORT SIZES.(U)
DEC 77 A HEDAYAT AFOSR-76-3050
UNCLASSIFIED AFOSR-TR-78-0487 NL

1 OF 1
ADA
052596

END
DATE
FILMED
5 -78
DDC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

2

6 Sampling Designs with Reduced Support Sizes.

9 Interim rept.,

by

10 A./Hedayat

Department of Mathematics
University of Illinois, Chicago

11 Dec [redacted] 77

12 27p.

18 AFOSR

19 TR-78-Ø487

---

1. Invited paper presented at International Conference on Optimization in Statistics held in IIT, Bombay, India during December 20-22, 1977.

2. Research supported by the Air Force Office of Scientific Research under Grant AFOSR-76-3Ø5Ø

15

16 23Ø4

17 A5

409 950 — mt

Sampling Designs with Reduced Support Sizes[1]

by

A. Hedayat[2]
Department of Mathematics
University of Illinois, Chicago

0. Introduction.

Suppose $\mathcal{U}$ is a finite population of N distinct units. Let $\mathcal{S}$ be the set of all subsets based on the elements of $\mathcal{U}$. A sampling design, d, based on $\mathcal{U}$ is a pair (S_d, P_d), where S_d is a subset of $\mathcal{S}$ and $P_d = \{p_d(s), s \in S_d\}$ is a probability distribution on S_d. To guarantee the estimability of the basic parameters of $\mathcal{U}$, such as the population total, we insist that the union of the subsets in S_d be $\mathcal{U}$ and $p_d(s) > 0$ for each s in S_d.

The first order inclusion probability of the unit i under d is defined to be

$$\pi_d(i) = \sum_{s \ni i} p_d(s),$$

and the second order (joint) inclusion probability of the units i and j $(i \neq j)$ under d is

$$\pi_d(i,j) = \sum_{s \ni i,j} p_d(s).$$

---

1. Invited paper presented at International Conference on Optimization in Statistics held in IIT, Bombay, India, during December 20-22, 1977.

2. Research supported by Grant AFOSR 76-3050A.

Since the introduction of unequal probability sampling by Horvitz and Thompson (1952), the emphasis in the theory has been towards working with the above inclusion probabilities. This paper is mainly concerned with problems related to these inclusion probabilities.

Two sampling designs d_1 and d_2 are said to be equivalent with respect to these inclusion probabilities if:

$$\pi_{d_1}(i) = \pi_{d_2}(i), \ \pi_{d_1}(i,j) = \pi_{d_2}(i,j), \quad \forall i,j.$$

This paper studies the extent to which these inclusion probabilities characterize the sampling designs. This study has led us to sampling designs which have applications in controlled sampling. We have studied the following problem, among others. The classical simple random sample of size n based on $\mathcal{U}$, denoted $SRS(N,n)$, is a sampling design whose support, S_d, consists of all $\binom{N}{n}$ possible samples of size n and whose probability distribution, P_d, is uniform on the support. Thus a problem of interest is to find sampling designs equivalent to $SRS(N,n)$ but whose support sizes may be less than $\binom{N}{n}$ and for which the probability distribution on their supports may or may not be uniform. It is shown that this can always be done. Such sampling designs have applications to controlled sampling.

1. Preliminaries

Let $\mathcal{U} = \{U_1, U_2, \ldots, U_N\}$ be a population of N identifiable units. Let $\mathcal{S}$ be the power set of $\mathcal{U}$, i.e., the set of all subsets based on the elements of $\mathcal{U}$. Note that the cardinality (size) of $\mathcal{S}$ is 2^N. Hereafter we shall refer to the units in $\mathcal{U}$ by their indices. Thus the unit U_i will be referred to by "i".

Definition 1.1. A sampling design, d, based on $\mathcal{U}$ is a pair (S_d, P_d), where S_d is a subset of $\mathcal{S}$ and $P_d = \{p_d(s), s \in S_d\}$ is a probability distribution on S_d. To guarantee the estimability of the basic parameters[(1)] of $\mathcal{U}$ we insist that the union of the subsets in S_d be $\mathcal{U}$ and $p_d(s) > 0$ for each s in S_d.

Throughout the paper the cardinality of a set Z will be denoted by $c(z)$.

Definition 1.2. S_d is called the support of d and $c(S_d)$ is called the support size of d.

Definition 1.3. A sampling design is said to be a uniform sampling design if P_d is uniform on S_d.

---

(1) Such as the population total $Y = \sum_{i=1}^{N} Y_i$, where Y_i is the value of a real-valued function on the unit i.

Definition 1.4. We say a sampling design d is of size n if $c(s) = n$ for all s in S_d.

In the sequel we shall refer to the elements of S_d as samples and a sample in S_d selected by implementing P_d as a probability sample. Perhaps the most adopted sampling design in practice is a sampling design which is known as a simple random sample design of size n which can be defined under our notation as:

Definition 1.5. A sampling design (S_d, P_d) based on $\mathcal{U}$ of size N is said to be a simple random sample design of size n, SRS(N, n), if

(i) S_d consists of all $\binom{N}{n}$ subsets of size n based on $\mathcal{U}$,

(ii) P_d is uniform on S_d, i.e., $p(s) = 1/\binom{N}{n}$.

In this paper we shall deal with sampling designs whose first order and second order inclusion probabilities are identical to the corresponding probabilities of simple random sample designs but whose support sizes may be less than $\binom{N}{n}$ and for which P_d may or may not be uniform. Such sampling designs have applications in the area of controlled sampling. The first order and second order inclusion probabilities are studied in Section 2.

2. Inclusion probabilities.

Since the introduction of unequal probability sampling by Horvitz and Thompson (1952) the emphasis in the theory has been towards working with the first and second order inclusion probabilities associated with sampling designs. These probabilities are defined as:

The first order inclusion probability associated with the unit i in $\mathcal{U}$ under the sampling design $d = (S_d, P_d)$ is

$$\pi_d(i) = \sum_{s \ni i} p_d(s). \tag{1}$$

This is the probability of selecting the unit i if we implement the sampling design d.

The second order (joint) inclusion probability associated with the units i and j $(i \neq j)$ in $\mathcal{U}$ under d is

$$\pi_d(i,j) = \sum_{s \ni i,j} p_d(s). \tag{2}$$

This is the probability of simultaneously selecting the units i and j if we implement the sampling design d.

Some known linear constraints on the inclusion probabilities π_i's and π_{ij}'s are:

Proposition 2.1. Under the sampling design d

$$\sum_{i=1}^{N} \pi_d(i) = \sum_{s} c(s) p_d(s), \tag{3}$$

$$\sum_{i(\neq j)}^{N} \pi_d(i,j) = \sum_{s \ni j} [c(s)-1]p_d(s), \tag{4}$$

$$\sum_{i=1}^{N} \sum_{j(\neq i)}^{N} \pi_d(i,j) = \sum_{s} c(s)[c(s)-1]p_d(s). \tag{5}$$

Corollary 2.1. *If* d *is a sampling design of size* n *then*

$$\sum_{i=1}^{N} \pi_d(i) = n, \quad \sum_{i(\neq j)}^{N} \pi_d(i,j) = (n-1)\pi_d(j), \quad j = 1,2,\ldots,N$$

$$\sum_{i=1}^{N} \sum_{j(\neq i)}^{N} \pi_d(i,j) = n(n-1). \tag{6}$$

Thus there are $N + 1$ distinct linear constraints on $\pi_d(i)$'s and $\pi_d(i,j)$'s. If the samples in S_d are not identical in size, then the expected sample size under d is

$$\text{expected sample size} = \sum_{s} c(s)p_d(s)$$

which is precisely $\sum_{i=1}^{N} \pi_d(i)$ and thus it should not be surprising that when d is a sampling design of size n then

$$\sum_{i=1}^{N} \pi_d(i) = n$$

whether or not d is uniform on its support.

3. The problems and background.

Because of the importance of $\pi_d(s)$'s and $\pi_d(i,j)$'s in the theory of sampling it is interesting to investigate the extent to which these inclusion probabilities characterize the sampling designs. For example, $\pi_d(i) = n/N$ and $\pi_d(i,j) = n(n-1)/N(N-1)$ if the sampling design is SRS(N,n). Then is it true that SRS(N,n) is the only design with $\pi_d(i) = n/N$ and $\pi_d(i,j) = n(n-1)/N(N-1)$? The answer is no. Indeed, such sampling designs exist which violate one or both conditions of SRS(N,n) specified in Definition 1.5. To formalize our problems we need the following definition.

Definition 3.1. Two sampling designs d_1 and d_2 based on $\mathcal{U}$ are said to be equivalent with respect to the first order and second order inclusion probabilities if

$$\pi_{d_1}(i) = \pi_{d_2}(i) \quad \text{and} \quad \pi_{d_1}(i,j) = \pi_{d_2}(i,j) \tag{7}$$

Hereafter, for simplicity, we shall say two sampling designs d_1 and d_2 are equivalent (designated by $d_1 \approx d_2$) if they are equivalent in the sense of Definition 3.1. Note that the condition $\pi_{d_1}(i) = \pi_{d_2}(i)$ implies that in order $d_1 \approx d_2$ it is necessary that the expected sample size under d_1 should be equal to the expected sample size under d_2, a natural demand for the concept to be practically meaningful.

Problem 1. Given a sampling design d_1, what is the minimum support size of a sampling design d_2 equivalent to d_1?

Problem 2. Suppose we are given a sampling design d_1 and a sampling design d_2 whose support size is minimum and is equivalent to d_1. Let M_{d_1} and M_{d_2} be the support size of d_1 and d_2 respectively. Then for what value of M, $M_{d_2} < M < M_{d_1}$, is there a sampling design with support size M which is equivalent to d_1?

These problems have not been fully solved as of to-day. Our experience indicated that they will remain unsolved for many years to come. Solutions to some aspects of these problems have been obtained by Chakrabarti (1963), Wynn (1977), Foody and Hedayat (1977), Hedayat and Li (1977) and Hedayat and Rao (1978).

Chakrabarti (1963) noticed that one can relate a balanced incomplete block (BIB) design based on N treatments in b blocks of size n to a sampling design by considering the blocks as samples and treatments as units and letting $p(s) = 1/b$ where s is a block of the design. Thus he proved that

Theorem 3.1. *A uniform sampling design of size n based on a population of size N is equivalent to SRS(N,n) if and only if it is associated with a BIB design with no repeated*

<u>blocks on</u> N <u>treatments in blocks of size</u> n.

Chakrabarti (1963) did not give any practical applications of sampling designs with support size less than $\binom{N}{n}$ and equivalent to SRS(N,n). Perhaps due to this lack of practical motivation of the problem solved by Chakrabarti, no further works on this subject came to print for a decade.

Avadhani and Sukhatme (1973) discussed sampling designs associated with BIB designs and gave some meaningful practical applications of such designs in controlled sampling. For actual examples of controlled sampling see, for example, Goodman and Kish (1950) and Avadhani and Sukhatme (1973).

A more systematic study of Problem 1 was done by Wynn (1977), who used Caratheodory's theorem [see Rockafellar (1970), p. 151] and, for example, proved that

<u>Theorem 3.2</u>. <u>If</u> d_1 <u>is a sampling design of size</u> n <u>based on a population of size</u> N <u>then there is a sampling design</u> $d_2 \approx d_1$ <u>with support size no greater than</u> $N(N-1)/2$.

For example, if d_1 is SRS(8,3) then there is a sampling design $d_2 \approx d_1$ whose support size is no greater than $8(7)/2 = 28$. For $N = 8$ and $n = 3$ the lower bound 28 is not sharp. Wynn (1977) gave an example of a sampling design with support size 24 equivalent to SRS(8,3).

Foody and Hedayat (1977) formalized the concept of **sampling design** of size n based on a **population of size** N equivalent to SRS(N,n) in the language of matrix algebra and mathematical programming and obtained several results in the terminology of BIB designs with repeated blocks. To point out some of their results and present further work in the area we need some notation and definitions which are given in Section 4. In the rest of the paper we shall limit our study to sampling designs of size n.

4. Sampling designs in the language of matrix algebra and mathematical programming.

A 2-element subset of $\mathfrak{u}$ of size N will be called a pair and an n-element subset will be called a sample of size n. Let P denote the incidence matrix (do not confuse with P_d) of pairs versus blocks. So P is a $\binom{N}{2}$ by $\binom{N}{n}$ zero-one matrix. Order the $\binom{N}{n}$ samples of size n in some fashion and let D be a multiset (a set which allow the elements to appear with multiplicity) based on $\binom{N}{n}$ samples of size n. We write f_i for the frequency of the ith sample of size n in D. Let $F_D \equiv (f_1,f_2,\ldots,f_{\binom{N}{n}})'$. Foody and Hedayat (1977) proved that

<u>Theorem 4.1</u>. <u>A frequency vector</u> F_D <u>determines a sampling design equivalent to</u> SRS(N,n) <u>if and only if</u>

$$PF_D = \lambda \underline{1} \tag{8}$$

<u>where</u> λ <u>is a positive integer and</u> $\underline{1}$ <u>is a column vector of all ones</u>.

<u>Proof</u>. The sampling design, d, associated with F_D can be constructed as follows: Let S_d consist of those n-element subsets of $\mathcal{U}$ whose corresponding f's in F_D is not zero. The probability associated with a sample in S_d will be the corresponding f divided by Σf_k. Now by (8)

$$\pi_d(i) = \frac{n\Sigma f_k/N}{\Sigma f_k} = \frac{n}{N}$$

and

$$\pi_d(i,j) = \frac{\lambda}{\Sigma f_k} = \frac{n(n-1)\Sigma f_k/N(N-1)}{\Sigma f_k} = \frac{n(n-1)}{N(N-1)}$$

which shows that $d \approx SRS(N,n)$. The necessity part of the theorem can be similarly proved.

At this point we would like to point out that the sampling design d associated with F_D will have support size less than $\binom{N}{n}$ if one or more components of F_D are zero and d will be nonuniform if there exist $i \neq j$ such that $f_i \neq f_j$.

Example 4.1. Let $N = 8$ and $k = 3$. Then SRS(8,3) has support size $\binom{8}{3} = 56$. The probability of each sample is 1/56. Based on Theorem 3.1 we exhibit below a sampling design equivalent to SRS(8,3) which is nonuniform and has support size 22.

Sample	Probability	Sample	Probability
125	1/56	347	2/56
137	1/56	128	3/56
146	1/56	178	3/56
245	1/56	268	3/56
246	1/56	468	3/56
367	1/56	478	3/56
467	1/56	234	4/56
127	2/56	567	4/56
237	2/56	136	5/56
256	2/56	145	5/56
257	2/56	358	6/56

Clearly the above is a sampling design and the reader can check for himself that for this design

$$\pi_d(i) = \frac{n}{N} = \frac{3}{8} \quad \text{and} \quad \pi_d(i,j) = \frac{n(n-1)}{N(N-1)} = \frac{6}{56}$$

as in the case of SRS(8,3).

Table 1 in Foody and Hedayat (1977) provides BIB designs which can be converted to nonuniform sampling designs with all possible support sizes 22 to 55 when $N = 8$ and $n = 3$. In this case SRS(8,3) is the only uniform sampling design.

Theorem 3.1 says that each feasible solution of the system

$$PF_D = \lambda \underline{1}, \quad F_D \geq 0 \tag{9}$$

corresponds to a sampling design equivalent to SRS(N,n). The set of all rational feasible solutions to this system corresponds to all sampling designs equivalent to SRS(N,n). Now there is always at least one rational feasible solution to (9), namely the solution corresponding to SRS(N,n) with $f_i = 1$ and $\lambda = \binom{N-2}{n-2}$. Using the language of mathematical programming we know that all feasible solutions to (9) are convex combinations of the basic feasible solutions, so the search for all sampling designs equivalent to SRS(N,n) reduces to finding all basic feasible solutions to (9); that is, to finding all of the vertices of the polytope defined by (9).

In practice we are not, of course, interested in finding all solutions to (9). Rather, we seek a solution which excludes, or at least minimizes the selection probability of certain samples. We may find such a sampling design by introducing an objective function which assigns positive cost to the samples which we wish to avoid and zero cost to the other samples. The standard linear programming algorithms for minimizing this objective function will then produce the desired design.

5. Bounds on the support size of a sampling design equivalent to SRS(N,n).

Let d^* be a sampling design whose support size, M_{d^*} is minimum among all sampling designs equivalent to SRS(N,n). Then we have

$$\left\{\frac{N}{n}\left\{\frac{N-1}{n-1}\right\}\right\} \leq M_{d^*} \leq \frac{N(N-1)}{2} \tag{10}$$

where $\{x\}$ denotes the smallest integer greater than or equal to x. Though the upper bound is already stated in Theorem 3.2, we can prove it easily by representing d^* in it equivalent form

$$PF_{D^*} = \lambda^* 1 \; . \tag{11}$$

Now recall that P has an $\binom{N}{2}$ rows and $\binom{N}{n}$ columns. Since $\binom{N}{2} = N(N-1)/2 < \binom{N}{n}$ thus rank of P is at most $N(N-1)/2$ [indeed it is precisely $N(N-1)/2$ by Lemma 5.1 of Foody and Hedayat (1977)]. Therefore $\lambda^* 1$ can be expressed as a linear combination of at most $N(N-1)$ columns of P meaning that F_{D^*} has at most $N(N-1)/2$ nonzero components. Thus $M_{D^*} \leq N(N-1)/2$.

To prove the lower bound, note that $\pi_{d^*}(i,j) \geq 1$. Thus to cover all pairs, each element of the population must appear in at least $\{(N-1)/(n-1)\}$ distinct samples in the support of d^*. Now let m be the smallest number of samples of size n needed to cover $\binom{N}{2}$ pairs. Thus, the average number of distinct samples in which each element appears, mn/N, must be at least $\{(N-1)/(n-1)\}$. Thus $M_d^* \geq \{(N/n)\{(N-1)/(n-1)\}\}$.

There are infinitely many N's and n's for which sampling designs equivalent to SRS(N,n) exist and have support size $\{(N/n)\{(N-1)/(n-1)\}\}$. Therefore, the lower bound in (10) is sharp. As an example, let $N = 7$, and $n = 3$. Then we have:

Example 5.1. Below is a sampling design with the minimum support size which is equivalent to SRS(7,3).

Sample	Probability	Sample	Probability
124	1/7	561	1/7
235	1/7	672	1/7
346	1/7	713	1/7
457	1/7		

Note that in this case $\{(N/n)\{(N-1)/(n-1)\}\} = 7$.

There are N's and n's for which the lower bound in (10) is much too large. For example, when $N = 8$ and $n = 3$ the lower bound in (10) becomes 11. But from Foody and Hedayat (1977) and Pesotchinsky (1977) we know that in this case the minimum support size is 22. In Example 4.1 we exhibited such a sampling design.

The problem of the characterization of all N's and n's for which the lower bound in (10) is achievable remains unsolved and it is a hard problem indeed. It is interesting to study the properties of S_d and P_d if the support size of d is minimum, as given in (10). One thing which we can say is this: If such a sampling design exists and if $\lceil (N/n)\lceil (N-1)/(n-1)\rceil\rceil = N$ then d must be a uniform sampling design and, moreover, d exists if and only if there is a BIB design with N blocks of size n based on N treatments.

Note that $\lceil (N/n)\lceil (N-1)/(n-1)\rceil\rceil \geq N$ and thus there is no sampling design equivalent to $SRS(N,n)$ and having support size less than N. Another question of interest is: Is there any sampling design equivalent to $SRS(N,n)$ with support size $N + 1$? The following proposition answers this question.

Proposition 5.1. *There is no sampling design, d, of size n based on a population of size N with properties*: (i) $d \approx SRS(N,n)$ and (ii) $c(S_d) = N + 1$.

Assume to the contrary and let d be such a sampling design. It is not difficult to see that there is an integer δ such that $\delta p_d(s)$ is an integer for all s in the support of d. Then the samples in S_d together with $\delta p_d(s)$'s form a BIB design based on $N + 1$ distinct blocks. This contradicts Theorem 3.2 of van Lint and Ryser (1972) which says there is no BIB design with precisely $N + 1$ distinct blocks.

In regard to problems listed in Section 3, our experience indicates that if d_1 is a SRS(N,n) and if there exists a sampling design, d_2, equivalent to SRS(N,n) with support size N then there are integers between N (the support of d_2) and $\binom{N}{n}$ (the support of SRS(N,n)) for which there are no sampling designs with such support sizes and equivalent to SRS(N,n). The result in Proposition 5.1 gives one such integer for arbitrary N and n. Let us consider the case of $N = 7$ and $n = 3$. In this case there are no sampling designs equivalent to SRS(7.3) with support sizes 8, 9, 10, 12. Whether or not there is a sampling design with support size 16 is unknown to this writer. However, if M is an integer between 7 and 35, and $M \neq 8, 9, 10, 12, 16$, then there is a sampling design with support size M and equivalent to SRS(7,3), according to Hedayat and Li (1977).

If the minimum support size is not N then we know very little about the support sizes of the sampling designs equivalent to SRS(N,n). Whether or not the case of $N = 8$ and $n = 3$ indicates something is not clear to us. In this case the minimum possible support size is 22 and, as Foody and Hedayat (1977) have shown, for every integer

$22 \leq M \leq 55$ there is a sampling design with support size M and equivalent to SRS(8,3). Any such sampling design for $N = 8$ and $n = 3$ will be nonuniform.

In Section 6 we shall study the method of trade off which is a very useful technique for finding sampling designs with support size smaller than $\binom{N}{n}$ and equivalent to SRS(N,n).

6. The method of trade off and its application in sampling.

The idea of trade off is as follows: For given N and n we shall write down SRS(N,n). Then in order to reduce the support of SRS(N,n) we shall try to find two sets of samples, S_1 and S_2, in the support of SRS(N,n) such that it is possible to remove S_2 from the support and assign the related probabilities to samples in S_1 in such a fashion that the resulting sampling design is equivalent to SRS(N,n). If this can be done, then we say S_2 has been traded off for S_1. But the theory which will be presented is much broader than this.

Recall the notation of Section 4 and let T be a nonzero integer column vector of dimension $\binom{N}{n}$.

<u>Definition</u> <u>6.1</u>. The vector T is called a <u>trade</u> if

$$PT = 0, \tag{12}$$

and the sum of all positive entries in a trade is called its <u>volume</u>.

Ignore the entries of T which are zero and let $t_1, t_2, \ldots, t_g$ denote the positive components and $t'_1, t'_2, \ldots, t'_h$ denote the negative components. Thus

$$\Sigma t_i + \Sigma t'_i = 0 \ . \tag{13}$$

Be definition of P and the existence of T we can immediately identify two sets of samples of size n

$$S_1 = \{s_1, s_2, \ldots, s_g\}, \ S_2 = \{s'_1, s'_2, \ldots, s'_h\}, \ S_1 \cap S_2 = \phi,$$

such that if (x,y) is a pair of elements in some sample of S_1 then

$$\sum_{s_i \ni (x,y)} t_i + \sum_{s'_i \ni (x,y)} t'_i = 0. \tag{14}$$

Conversely, if we are given two sets of samples and two sets of integers of the form:

sample	integer	sample	integer	
s_1	t_1	s'_1	t'_1	
s_2	t_2	s'_2	t'_2	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$s_i \neq s'_j$
s_g	t_g	s'_h	t'_h	

with properties (13) and (14) we can immediately write down a vector T which is a trade. The following example elucidates the above argument.

Example 6.1. Let $N = 7$ and $n = 3$. Consider the following samples and related integers

sample	integer	sample	integer
123	2	124	-1
145	1	125	-1
156	1	135	-1
246	1	136	-1
257	1	236	-1
356	1	237	-1
367	1	456	-1
		567	-1

The reader can check for himself that these two sets of samples and corresponding integers satisfy (13) and (14). As an example let $(xy) = (12)$ we see that the sample 123 contains (12) with $t = 2$. In the set of samples with negative integers there are two samples 124 and 125 with $t'_1 = -1$ and $t'_2 = -1$.

To write the corresponding vector T associated with Example 6.1, let T be a column vector of size $\binom{7}{3} = 35$. Note that by definition of P each component of T is related to a specific sample. For those samples listed above enter their corresponding t's or t''s in the appropriate components of T and enter zero for all other samples.

For given N and n let F_D be a frequency vector as defined in Section 4. Then we have:

Theorem 6.1. *If T is a trade and if F_D determines a sampling design equivalent to $SRS(N,n)$ then $F_D + T$ determines a sampling design which is equivalent to $SRS(N,n)$ provided that no entry of $F_D + T$ is negative.*

Example 6.2. Let $N = 7$ and $n = 3$ and let F_D be the column vector with all its entries equal to 1. Let T be the trade exhibited in Example 6.1. Then $F_D + T$ provides us a sampling design which is equivalent to $SRS(7,3)$. Note that the support size of the corresponding design is 27 and the design will be nonuniform. The corresponding sampling design can be easily obtained as follows. Delete from the support $SRS(7,3)$ [note that the sampling design associated with our choice of F_D is precisely $SRS(7,3)$] those samples whose related integers are -1. Since there are 8 such samples we will be left with $35 - 8 = 27$ samples. These 27 samples will be the support of the sampling design associated with $F_D + T$. The corresponding probabilities are calculated as follows: The probability associated with a sample in the new support will remain the same if that sample did not appear in the trade. Otherwise, the corresponding new probability is $1/35 + t/35$, where t is its related integer in T. Thus in this case probability associated with the sample (123) will be 3/35.

Note that the trade in Example 6.1 can be used in finding sampling designs with support sizes smaller than $\binom{N}{3}$ and equivalent to SRS(N,3), as long as $N \geq 7$.

Hedayat and Li (1977) have studied the theory of trade off in the context of BIB designs with repeated blocks and have obtained several results directly applicable in sampling. For example, they have shown that:

Theorem 6.3. *A trade of volume* i *exists if and only if* $i \neq 1, 2, 3$ or 5.

Due to limitation of space, several other results on trade off will be reported elsewhere.

REFERENCES

Avadhani, M. S. and Sukhatme, B. V. (1973). Controlled sampling with equal probabilities and without replacement. *Internat. Statist. Rev.* **41** 175-182.

Chackrabarti, M. C. (1963). On the use of incidence matrices in sampling from finite populations. *J. Indian Statist. Assoc.* **1** 78-85.

Foody, W. and Hedayat, A. (1976). On theory and application of BIB designs with repeated blocks. *Ann. Statist.* **5** 932-945.

Goodman, R. and Kish, L. (1950). Controlled selection - A technique in probability sampling. J. Amer. Statist. Assoc. 45 350-372.

Hedayat, A. and Li, Shuo-Yen, R. (1977). On theory of trade off and its application to the construction of BIB designs with variable support sizes. Technical report, Department of Mathematics, University of Illinois, Chicago.

Hedayat, A. and Rao, J. N. K. (1978). Contributions to finite population samplings. To appear.

Horvitz, D. G. and Thompson, D. J. (1952). A generalisation of sampling without replacement. J. Amer. Statist. Assoc. 47 663-685.

Pesotchinsky, L. (1977). Remarks on BIB designs with repeated blocks. Technical report, Department of Statistics, Stanford University.

Rockafellar, R. T. (1970). Convex Analysis. Princeton Univ. Press.

van Lint, J. H. and Ryser, H. J. (1972). Block designs with repeated blocks. Discrete Math. 3 381-396.

Wynn, H. P. (1977). Convex sets of finite population plans. Ann. Statist. 5 414-418.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 78-0487	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SAMPLING DESIGNS WITH REDUCED SUPPORT SIZES		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A. Hedayat		8. CONTRACT OR GRANT NUMBER(s) AFOSR 76-3050
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Illinois Department of Mathematics Box 4348, Chicago, IL 60680		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, DC 20332		12. REPORT DATE December 1977
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

controlled sampling, efficient sampling, support of a sampling design

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Suppose $\mathcal{U}$ is a finite population of N distinct units. Let $\mathcal{S}$ be the set of all subsets based on the elements of . A sampling design, d, based on $\mathcal{U}$ is a pair (S_d, P_d), where S_d is a subset of $\mathcal{S}$ and $P_d = \{p_d(s), s \in S_d\}$ is a probability distribution on S_d. To guarantee the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20 . Abstract

estimability of the basic parameters of $\mathcal{U}$, such as the population total, we insist that the union of the subsets in S_d be $\mathcal{U}$ and $P_d(s) > 0$ for each s in S_d.

The first order inclusion probability of the unit i under d is defined to be

$$\Pi_d(i) = \sum_{s \ni i} p_d(s),$$

and the second order (joint) inclusion probability of the units i and j $(i \neq j)$ under d is

$$\Pi_d(i,j) = \sum_{s \ni i,j} p_d(s).$$

Since the introduction of unequal probability sampling by Horvitz and Thompson (1952), the emphasis in the theory has been towards working with the above inclusion probabilities. This paper is mainly concerned with problems related to these inclusion probabilities.

Two sampling designs d_1 and d_2 are said to be equivalent with respect to these inclusion probabilities if:

$$\Pi_{d_1}(i) = \Pi_{d_2}(i), \quad \Pi_{d_1}(i,j) = \Pi_{d_2}(i,j), \quad \forall k,j.$$

This paper studies the extent to which these inclusion probabilities characterize the sampling designs. This study has led us to sampling designs which have applications in controlled sampling. We have studied the following problem, among others. The classical simple random sample of size n based on $\mathcal{U}$, denoted SRS(N,n), is a sampling design whose support, S_d, consists of all $\binom{N}{n}$ possible samples of size n and whose probability distribution, p_d, is uniform on the support. Thus a problem of interest is to find sampling designs equivalent to SRS(N,n) but whose support sizes may be less than $\binom{N}{n}$ and for which the probability distribution on their supports may or may not be uniform. It is shown that this can always be done. Such sampling designs have applications to controlled sampling.